Application of Adverse Outcome Pathways to U.S. EPA's Endocrine Disruptor Screening Program

Patience Browne,[1] Pamela D. Noyes,[1] Warren M. Casey,[2] and David J. Dix[1]

[1]Office of Chemical Safety and Pollution Prevention, U.S. Environmental Protection Agency, Washington, DC, USA
[2]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA

BACKGROUND: The U.S. EPA's Endocrine Disruptor Screening Program (EDSP) screens and tests environmental chemicals for potential effects in estrogen, androgen, and thyroid hormone pathways, and it is one of the only regulatory programs designed around chemical mode of action.

OBJECTIVES: This review describes the EDSP's use of adverse outcome pathway (AOP) and toxicity pathway frameworks to organize and integrate diverse biological data for evaluating the endocrine activity of chemicals. Using these frameworks helps to establish biologically plausible links between endocrine mechanisms and apical responses when those end points are not measured in the same assay.

RESULTS: Pathway frameworks can facilitate a weight of evidence determination of a chemical's potential endocrine activity, identify data gaps, aid study design, direct assay development, and guide testing strategies. Pathway frameworks also can be used to evaluate the performance of computational approaches as alternatives for low-throughput and animal-based assays and predict downstream key events. In cases where computational methods can be validated based on performance, they may be considered as alternatives to specific assays or end points.

CONCLUSIONS: A variety of biological systems affect apical end points used in regulatory risk assessments, and without mechanistic data, an endocrine mode of action cannot be determined. Because the EDSP was designed to consider mode of action, toxicity pathway and AOP concepts are a natural fit. Pathway frameworks have diverse applications to endocrine screening and testing. An estrogen pathway example is presented, and similar approaches are being used to evaluate alternative methods and develop predictive models for androgen and thyroid pathways. https://doi.org/10.1289/EHP1304

Introduction

Many chemicals have the potential to interfere with normal endocrine functioning, which may lead to a variety of adverse outcomes including developmental deformities, impaired reproduction, and decreased survival. Potential adverse outcomes following exposure to endocrine-active substances have been the subject of intensive study and have been described in numerous research papers and reviews (e.g., Colborn and Clement 1992; Kavlock et al. 1996; WHO 2002; WHO/UNEP 2012; Hotchkiss et al. 2008; Soto and Sonnenschein 2010; Nohynek et al. 2013; Gore et al. 2015a). Although most research studies focus on one endocrine pathway or on one part of one endocrine pathway, the endocrine system is inherently integrative and adaptive. Endocrine effects can vary enormously by the organ and time point examined, even within the same individual. Conclusions from various researchers or reviewers on endocrine disruption have sometimes been divergent or even contradictory, suggesting that our scientific understanding of the etiology of adverse outcomes in humans and wildlife through endocrine mechanisms is far from complete. Many organizations including the U.S. EPA, the National Institutes of Health (NIH), the Organisation for Economic Co-operation and Development (OECD), the World Health Organization (WHO), and the United Nations Environmental Programme (UNEP) have supported research, developed guidance, and published standardized test guidelines to evaluate endocrine disruption in humans and wildlife.

U.S. EPA's Endocrine Disruptor Screening Program

The U.S. EPA's Endocrine Disruptor Screening Program (EDSP) screens and tests chemicals to determine potential endocrine effects in humans and wildlife. The EDSP was established in 1998 following amendments to the Federal Food, Drug, and Cosmetic Act (FFDCA) and the Safe Drinking Water Act (SDWA), mandating the U.S. EPA to screen chemicals for potential estrogenic effects in humans and providing authority to include other endocrine effects (U.S. Congress 1996a, 1996b). In response, the U.S. EPA convened the Endocrine Disruption Screening and Testing Advisory Committee (EDSTAC), consisting of regulatory, industry, and academic experts, to advise the agency on development and implementation of an endocrine disruptor screening program. The committee recommended expanding the scope of the EDSP to evaluate chemical effects on the androgen and thyroid pathways in wildlife and humans and to do so employing a two-tiered screening and testing strategy (EDSTAC 1998).

The first tier of assays screens chemicals for potential activity in estrogen, androgen, and thyroid pathways in both sexes of several vertebrate taxa. The battery of 11 complementary assays includes five *in vitro* assays that provide mechanistic data and six short-term, *in vivo* assays including bioassays measuring changes in organ weights and assays conducted in organisms with functional neuroendocrine axes (Figure 1). Tier 1 assays were designed to maximize sensitivity; however, considering collective results from multiple complementary assays relevant to each endocrine pathway was intended to reduce the limitations of each individual assay and to provide confidence in the hypothesized mode of action (U.S. EPA 2011).

Results of the Tier 1 screening battery were considered with other scientifically relevant information (OSRI; e.g., guideline studies submitted to the U.S. EPA in the pesticide registration process, research published in peer-reviewed literature; https://www.epa.gov/endocrine-disruption/status-endocrine-disruptor-screening-program-other-scientifically-relevant) to determine the weight

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

of evidence (WoE) supporting a chemical's potential endocrine activity. Criteria considered in WoE evaluations were described in a guidance document (U.S. EPA 2011) and included consideration of the nature of effects within and across studies and their biological plausibility, consistency of biological effects observed within and among species/sexes both within Tier 1 assays and OSRI, and if effects occurred in the absence of systemic toxicity. Integrating data from multiple studies conducted at various levels of biological organization to arrive at a determination of "potential endocrine activity" (or the absence thereof) can present a challenge for interpretation. To facilitate the collective interpretation of multiple studies, EDSP Tier 1 screening data were conceptually organized around hypothesized modes of action (e.g., altered receptor signaling, altered hormone synthesis, altered neuroendocrine axis function) in "estrogenic," "antiestrogenic," "androgenic," "antiandrogenic," and "thyroid-active" pathways (EDSTAC 1998; U.S. EPA 2011).

For chemicals determined to have potential endocrine activity, four longer-term, Tier 2 tests conducted in mammals, fish, amphibians, and birds may be requested to characterize dose–response relationships and adverse effects. Tier 2 tests include apical end points necessary for risk assessment that are regulated by endocrine and nonendocrine biological pathways, such as changes in growth, development, and reproduction (Figure 1). Together, the EDSP screening and testing strategy links mechanistic data to apical end points and is a unique regulatory program designed around a toxicological mode of action framework (EDSTAC 1998; U.S. EPA 2011).

The EDSP chemical universe comprises approximately 10,000 unique chemicals, including pesticide active and nonpesticide inert ingredients, as well as a large number of contaminants known or anticipated to occur in drinking water, such as industrial chemicals, pharmaceuticals, and disinfection byproducts (U.S. EPA 2012). The chemical domain includes both data-rich chemicals subject to substantial *in vivo* testing prior to use (e.g., pesticide active ingredients) and data-poor chemicals with limited toxicological or use information (e.g., nonpesticide industrial chemicals). The first Tier 1 test orders for 58 pesticide active and 9 pesticide inert ingredients were issued in 2009 (U.S. EPA 2009). The manufacturers of eight active and seven inert chemicals voluntarily opted out of the pesticide market, and data for the remaining 52 "List 1 chemicals" were submitted to the U.S. EPA. The resulting WoE determinations of potential endocrine activity and identification of additional data needed to definitively determine chemical effects were finalized in 2015 (U.S. EPA 2015a). A second list of 107 chemicals was published in 2013 (https://www.epa.gov/endocrine-disruption/overview-second-list-chemicals-tier-1-screening-under-endocrine-disruptor), but test orders have yet to be issued. Based on the timeline to date, screening all chemicals in the EDSP universe using the EDSP Tier 1 battery would require decades, many millions of dollars, and large numbers of laboratory test animals. Alternatively, the availability of *in vitro* high-throughput screening (HTS) and computational data for thousands of chemicals provides an information source to more efficiently prioritize chemicals for further evaluation based on indications of potential endocrine activity and may make list-driven screening of relatively few chemicals obsolete.

Use of HTS and Computational Toxicology Tools in the EDSP

When the EDSP was initially conceived in 1998, *in vitro* HTS assays were proposed as an initial method of providing
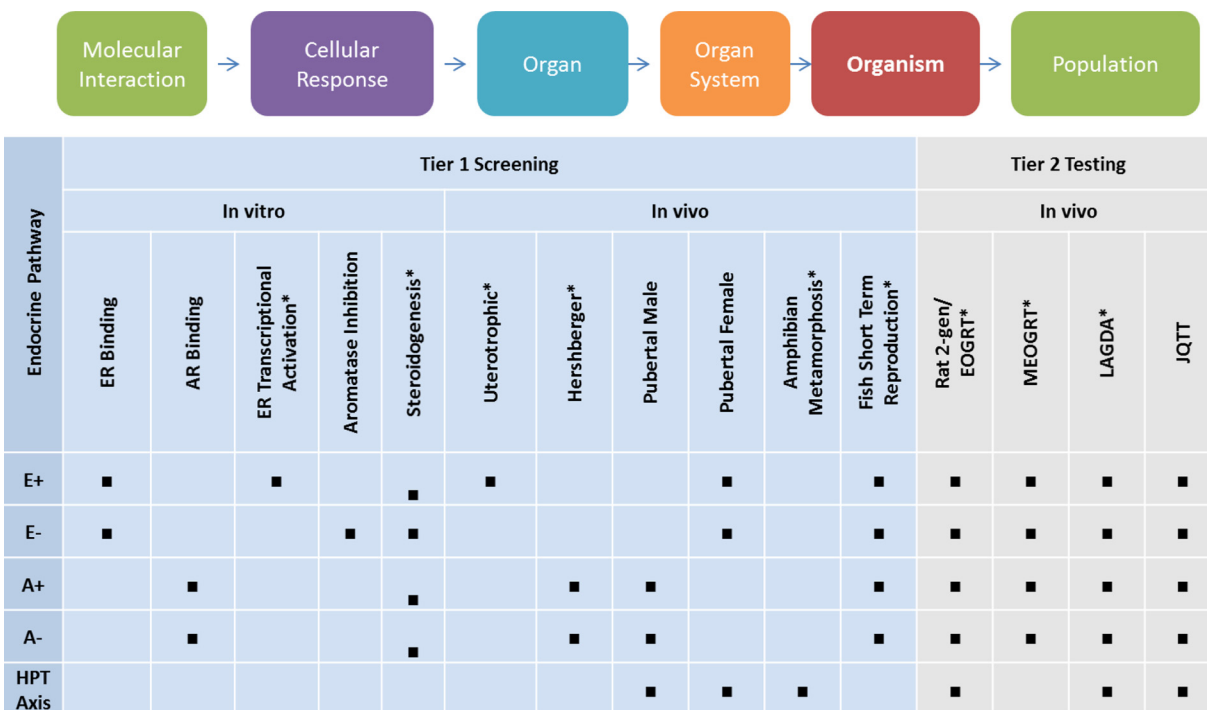
Figure 1. U.S. EPA Endocrine Disruptor Screening Program (EDSP) battery of 11 Tier 1 screening assays for activity and Tier 2 tests for identifying dose–response relationships and adverse effects[a]. Screening and testing data are interpreted for each endocrine pathway, although intact animal *in vivo* responses may involve multiple end points and pathways. Levels of biological complexity from molecular interactions through to populations are represented by the Tier 1 and Tier 2 screens and tests, consistent with an adverse outcome pathway (AOP) framework. A+, androgenic; A–, antiandrogenic; E+, estrogenic; E–, antiestrogenic; HPT axis, hypothalamic–pituitary–thyroid axis. For more detail about specific test methods and protocols, refer to EDSP test guidelines (https://www.epa.gov/test-guidelines-pesticides-and-toxic-substances/series-890-endocrine-disruptor-screening-program).
[a]EPA test guidelines harmonized through the OECD.

mechanistic data and prioritizing chemicals for further screening (EDSTAC 1998). At the time, the availability and reliability of commercial *in vitro* assays were limited. In the subsequent years, major technological advances have produced abundant HTS tools with applications for toxicity testing. Federal programs such as the Tox21 collaboration (http://www.ncats.nih.gov/tox21) and the U.S. EPA's ToxCast™ program (http://www2.epa.gov/chemical-research/toxicity-forecasting) use *in vitro* HTS assays to screen thousands of chemicals across hundreds of molecular targets and include assays relevant to estrogen, androgen, and thyroid pathway signaling.

These HTS tools have clear utility in the EDSP program, can increase the rate of chemical screening, and can identify chemicals likely to be the most biologically active in humans and wildlife. Consequently, the EDSP is now incorporating HTS data in the endocrine screening and testing framework (Browne et al. 2015; U.S. EPA 2015b). Integrating the results of high-throughput mechanistic data with the Tier 1 screening battery and other relevant information is aided by both the data organization framework for WoE screening evaluations and the EDSP's design around hypothesized endocrine modes of action.

Toxicity Pathways and Adverse Outcome Pathways

Toxicity pathways, as described in the National Research Council (NRC) report *Toxicity Testing in the 21st Century* (NRC 2007), are cellular response pathways that when sufficiently perturbed result in adverse health effects but do not necessarily include a molecular initiating event (MIE) or an adverse outcome (OECD 2012). Adverse outcome pathways (AOPs) represent an evolution of the toxicity pathway concept and describe a framework for linking the mechanism of chemical interaction with the apical end points used for risk assessment and regulatory decision making (Ankley et al. 2010). The concepts underlying toxicity pathways and AOPs are not new, and similar approaches for relating mechanistic interactions to downstream biological events have been described in constructs such as "mechanisms of action" and "modes of action." Although these concepts share similarities, the various terms have contributed to substantial diversity in definitions of and components included in toxicity pathways (Whelan and Andersen 2013). Recent efforts have attempted to avoid similar confusion by developing precise vocabulary and by defining criteria for evaluating candidate AOPs (Villeneuve et al. 2014a).

AOPs begin with an MIE and culminate in an adverse outcome linked by a series of biologically plausible and measurable intermediate key events at increasingly complex levels of biology from molecular responses to cellular and organ system perturbations. Relationships between key events may be causal, inferential, or putative and may be based on *in vitro*, *in vivo*, or computational data. AOPs were initially developed for ecotoxicology, in which an adverse outcome in an individual can be plausibly linked to population-level effects (Ankley et al. 2010; Kramer et al. 2011). More recently, AOPs have been adopted for human health assessment, in which case adversity is considered a detrimental effect observed in the individual (Patlewicz et al. 2015). For the purposes of this discussion, we will consider "toxicity pathways" to be a part of one or more potential AOPs (Figure 2). Although both toxicity pathways and AOPs are simplifications of complex biological processes, they provide systematic organizing frameworks to link mechanistic information to data collected over different biological scales and evaluate underlying biology knowledge (or gaps therein). It should be noted that AOPs are primarily tools for characterizing hazard. A variety of factors play a role in the exposure-to-outcome continuum that may alter the biological effects of a chemical. Pharmacokinetic studies and *in silico* models can improve predictivity by linking

exposure to the toxicological effect (Teeguarden et al. 2016), but in both cases, these are considered outside of the scope of AOPs (NAS 2017).

To support AOP development and foster collaboration and coordination among an international community, an AOP Knowledge Base was developed by the OECD, the U.S. EPA, the U.S. Army Corps of Engineers, the European Commission Joint Research Centre, and other partners (http://aopkb.org/). In addition to functioning as a repository of AOP information, the AOP Knowledge Base is also expected to promote collective participation of a broader scientific and regulatory community in AOP development, evaluation, exploration, and application. Once an AOP is described, the empirical evidence and the strength of predictive relationships between key events and adverse outcomes can be evaluated using modified Bradford-Hill criteria to assess the strength of the experimental methods and the biological relevance of the observed responses (Vinken 2013; Villeneuve et al. 2014b; Becker et al. 2015; http://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm).

A single MIE (e.g., a ligand binding to the estrogen receptor) may be associated with many separate AOPs, and similarly, an adverse outcome (e.g., reduced fecundity) may result from the perturbations in any one of many separate pathways. The number and specificity of intermediate key events included and the acceptable level of uncertainty in the AOP may vary with the intended application. Development of detailed individual AOPs may provide valuable insights into underlying toxicological and physiological processes, but such fine-scale consideration of biological pathways is not always needed in regulatory science. To better simulate the complexity of biological systems and for application to chemical regulation, multiple AOPs can be used to build AOP networks that better approximate biology (Villeneuve et al. 2014a). AOP networks integrate several MIEs or adverse outcomes, or both, that share at least one common element (Knapen et al. 2015; Villeneuve et al. 2014b). Networks that share intermediate key events can identify points of biological convergence common to more than one pathway.

Objectives

This paper discusses potential applications of AOPs and toxicity pathways to endocrine screening and testing as an organizational tool for integrating HTS assays and computational toxicology with traditional guideline toxicological studies used for making regulatory decisions. Specifically, we describe the application of these tools to organize endocrine data for WoE evaluations of a chemical's endocrine activity potential. Because WoE considers end points that are not measured in the same assay, identifying plausible linkages of an endocrine mode of action to downstream effects is needed to identify chemicals as endocrine disruptors and to examine the consistency of biological responses across independent studies. AOPs and toxicity pathways can help to elucidate the taxonomic conservation of endocrine responses, and thus the relevance of mammalian toxicology data (for human health safety assessment), to ecotoxicology in nonmammalian wildlife and vice versa. Using this same organizational framework, we can evaluate the ability of computational tools measuring an MIE or a key event to predict downstream effects and to examine the utility of pathway frameworks for developing integrative approaches for endocrine testing. We also discuss potential applications of AOPs for future endocrine testing strategies. Although the endocrine AOPs discussed herein do not include all key events and may characterize complex AOP networks with extremely simplified biology, they represent the end points that are available for regulatory decision making in the U.S. EPA's EDSP. Organizing data using an AOP framework, regardless of how simplified it

may be, helps to examine the consistency of responses across independent assays and to build confidence around the hypothesized endocrine mode of action.

The EDSP evaluates the potential of environmental chemicals to interact with the estrogen, androgen, and thyroid systems, and it has a narrower toxicological focus than other U.S. EPA programs. This limited toxicological domain was used to demonstrate how innovative 21st century toxicology tools could be used in a regulatory context (Rotroff et al. 2013; Reif et al. 2010), and the endocrine program was an early adopter of these approaches (Browne et al. 2015; U.S. EPA 2015b). Other offices in the U.S. EPA including the Office of Pesticide Programs (LaLone et al. 2017; LaLone et al. 2013; https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/strategic-vision-adopting-21st-century-science), the Office of Pollution Prevention and Toxics (https://www.epa.gov/tsca-screening-tools), and the National Center for Environmental Assessment (Makris et al. 2016; U.S. EPA 2015c) are now considering toxicity pathway and AOP approaches for incorporating new technology in chemical hazard identification.

## Current Applications of Pathway Frameworks in the EDSP

Organizing Frameworks

The two-tiered approach used by the EDSP to evaluate the potential effects of environmental chemicals assumes underlying biological links between end points measured in different assays. Although the representation is overly simplistic because some *in vivo* assays include end points that account for responses at multiple levels of biological complexity, EDSP assay end points can be mapped to a generic AOP for each endocrine pathway evaluated by the program (Figure 1). Tier 1 *in vitro* screening assays measure possible endocrine MIEs and capture early cellular responses, whereas short-term Tier 1 *in vivo* bioassays provide whole-organism and organ system responses to chemical exposures. The Tier 1 screening battery is intended to show the potential for endocrine activity rather than to represent conclusive evidence of potential adverse outcomes (Figures 1 and 2). For this reason, the Tier 1 screening assays can be interpreted in endocrine toxicity pathways rather than in AOPs because these assays are not designed to measure long-term adverse responses in individuals or in populations. In contrast, Tier 2 assays include apical end points, such as impaired growth or reproduction at individual and population levels, but they do not include mechanistic data and are typically sensitive to more than one mode of action. As a result, there are no assurances that endocrine activity determined from screening assays causes an apical response measured in longer-term studies (Coady et al. 2017). However, AOPs provide a systematic approach for organizing available information and supporting inferential links between mechanisms and adverse effects (Wittwehr et al. 2017), and they help to create plausible, causal links between data derived from varied sources.

As mentioned above, the amendments to the FFDCA and the SDWA specifically mandate that the U.S. EPA examine effects of environmental chemicals as potential estrogen agonists, and for this reason, both EDSP assays and endocrine HTS assays have more coverage of the estrogen pathway than they have of other pathways considered by the U.S. EPA's endocrine program. When estrogen-relevant end points measured in EDSP Tier 1 and Tier 2 assays are organized in an AOP framework (Figures 3 and 4), the resulting AOP does not include all intermediate key events. Despite the missing components, organizing the end points in an AOP of "estrogenic responses" allows one to determine the consistency of the response within an assay and across multiple assays and contributes to the WoE (or the lack thereof) that the observed responses are due to chemicals interacting with the estrogen system. For example, an estrogen agonist MIE, as measured by *in vitro* estrogen receptor (ER) binding and estrogen receptor transactivation (ERTA) assays, can support a plausible mechanism for an observed increase in uterine weight of immature or ovariectomized female rats following chemical exposure and is assumed to be primarily mediated through ERα genomic signaling linked to cell proliferation and increased water imbibition (Figure 3). This endocrine mechanism may be further supported by changes in the weight and histology of reproductive organs, altered estrous cyclicity, and changes in the onset of reproductive maturity measured in the female rat pubertal assay. Results of the Tier 2 extended one-generation reproductive toxicity study (EOGRTS) in rodents may contribute additional support for a chemical's hypothesized estrogen agonist activity if chemical exposure is associated with altered organ-, organ system–, and organism-level responses (Figure 3). Although this is not a complete AOP in the sense that all key events are not clearly delineated, this more generalized approach to an AOP includes the end points that are currently included in the U.S. EPA's screening and testing program, and these are the data used for regulatory decision making to evaluate the potential estrogen agonist activity of a chemical.

A similar approach can be adopted using AOP frameworks to make biological linkages between screening and testing responses in an ecotoxicological context. Tier 1 mechanistic screening data can be linked to various end points in the fish short-term reproduction assay (FSTRA) and to apical responses in the Tier 2 Medaka extended one-generation reproduction test (MEOGRT) (Figure 4). A chemical binding and activating the ER may be linked to adverse effects such as reduced fertility and fecundity as well as declining population trajectories (Groh et al. 2015; Ankley et al. 2010). Similar to the manner in which Figure 3 indicates plausible relationships between mammalian end points, Figure 4 does not identify every key event in the estrogen agonist
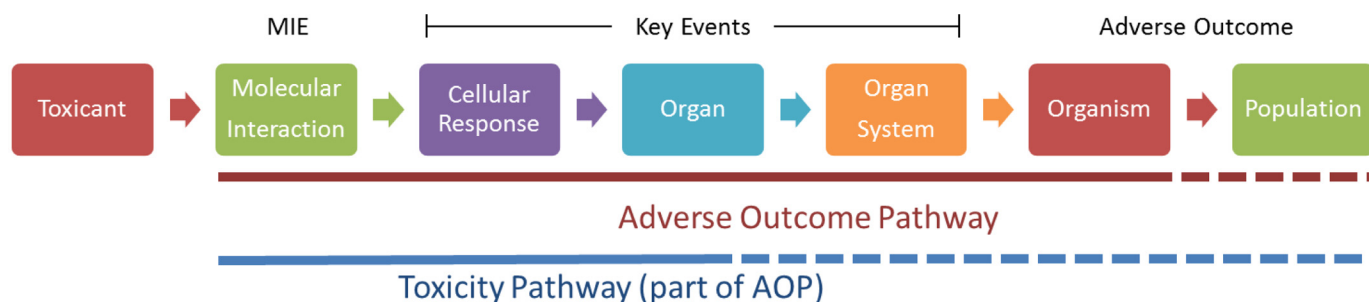
Figure 2. Adverse outcome pathways (AOPs) begin with a molecular interaction serving as a molecular initiating event (MIE), leading to a series of key events and eventually to an adverse outcome at the organismal level for human health, and at the population level for ecotoxicology assessments. Toxicity pathways can be considered part of an AOP, including plausibly linked molecular interactions and key events, but may not include an adverse outcome.
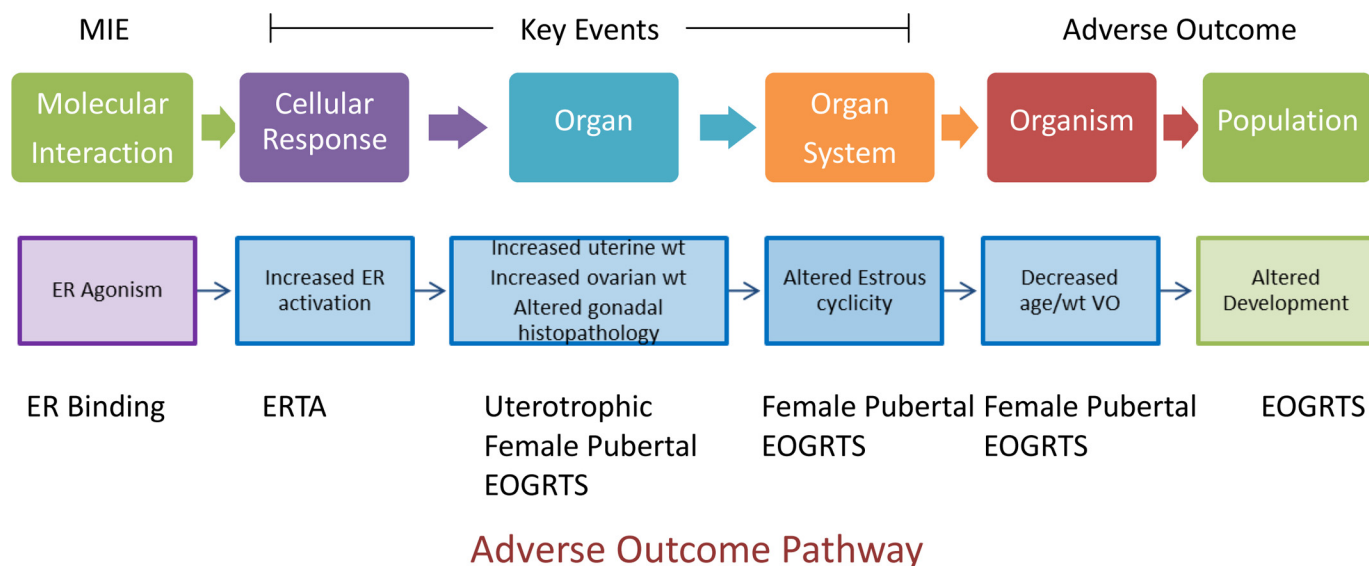
Figure 3. EDSP Tier 1 and Tier 2 end points relevant to the female mammalian estrogen agonist signaling pathway can be organized using an adverse outcome pathway (AOP) framework. Estrogen receptor (ER) binding and activation [i.e., the molecular initiating event (MIE)] can be linked to several related key events from EDSP Tier 1 and Tier 2 assays, leading to an adverse outcome (e.g., altered development). It should be noted that in this example, the AOP only includes regulatory end points included in U.S. EPA test guidelines used to evaluate the potential endocrine activity of environmental chemicals. EOGRTS, extended one-generation reproductive toxicity study; ERTA, estrogen receptor transactivation; VO, vaginal opening.

AOP for teleost fish, but it organizes end points that are currently used to make regulatory decisions about the potential estrogenicity of chemicals in a biologically grounded framework that supports interpretation of the screening and testing assays.

Although the effects of estrogen agonists in mammals and in fish differ, the ERα and the concomitant signaling pathway are highly conserved in vertebrates (Ankley et al. 2016), and both the rodent and fish assays were included in the EDSP screening battery to provide information on endocrine effects in all vertebrates (Figure 1). The U.S. EPA's guidance for WoE describes key lines of inquiry including agreement of outcomes within an individual assay (i.e., "complementarity") and among the different assays in the battery (i.e., "redundancy"; U.S. EPA 2011). Taken together, the use of AOPs and toxicity pathways to organize end points measured in EDSP guideline tests allows one to consider the effects across multiple AOPs linked to the same MIE (as in

Figure 1), helps to identify both complementarity and redundancy among assay data, aids in the interpretation of results from varied study types, and builds plausible links between mechanistic and apical responses (LaLone et al. 2016; LaLone et al. 2013). In addition, organizing EDSP assays and end points along an AOP framework can aid in understanding temporal relationships between key events and potential transient effects that may not necessarily lead to an adverse response (Ankley and Villeneuve 2015).

Systematic organization of data in AOPs and pathway frameworks also facilitates the inclusion of data from sources other than EDSP guidelines that can help to bridge gaps by providing information on intermediate key events not measured in Tier 1 and Tier 2 assays, further increasing confidence in an endocrine model of action leading to an adverse outcome. The EDSP now considers HTS data in the endocrine screening and testing
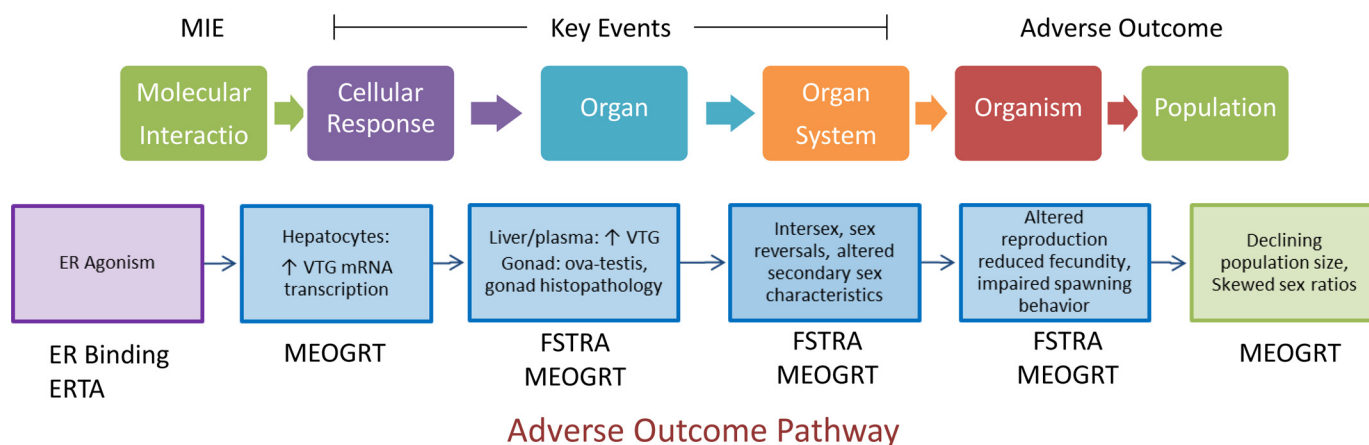
Figure 4. EDSP Tier 1 and Tier 2 test assays mapped to an ecotoxicological adverse outcome pathway (AOP) for an estrogen receptor (ER) agonist in male fish including a molecular initiating event (MIE) of receptor binding and related key events measured in Tier 1 and Tier 2 assays and terminating in an adverse outcome represented by declines in population size and by altered sex composition. FSTRA, fish short-term reproductive assay; MEOGRT, Medaka extended one-generation reproductive test; VTG, vitellogenin (egg precursor protein).

framework (U.S. EPA 2015b). As discussed above, when conceived in the late 1990s, the intention was to include HTS data in endocrine screening, and with the recent availability of HTS assay data from programs such as ToxCast™ and Tox21, MIEs and early key events for thousands of chemical structures can be elucidated. These data may contribute to the WoE or, in some instances, may obviate requirements for EDSP assays (Browne et al. 2015).

The ToxCast™ and Tox21 programs generate *in vitro* data that include end points covered in the EDSP Tier 1 screening battery. In some cases, these assays perform as well as or better than their low-throughput *in vitro* Tier 1 counterparts and cost substantially less; furthermore, the HTS test systems have the capacity to screen thousands of chemicals every year (e.g., Kleinstreuer et al. 2016b; Browne et al. 2015; Judson et al. 2015). In addition, the ToxCast™ and Tox21 HTS assays include endocrine targets that are not currently part of the EDSP screening and can expand the data used to evaluate potential endocrine effects of environmental chemicals (e.g., Filer et al. 2014, Reif et al. 2010). For example, when the Tier 1 battery was developed, reliable *in vitro* assays relevant to the thyroid hormone pathway were not identified. *In vitro* thyroid hormone receptor assays are now available, and several other thyroid pathway *in vitro* assays are now under development (Hallinger et al. 2017; Paul Friedman et al. 2016; OECD 2014). Mechanistic HTS data may replace current Tier 1 *in vitro* end points, expand the understanding of how chemicals interact with the endocrine system by providing information on additional molecular targets, increase the speed and efficacy of environmental chemical hazard evaluation, and aid in predicting the outcome of whole-animal assays when integrated with other evidence. AOPs provide a roadmap for evaluating these HTS data with traditional *in vivo* toxicology end points and for examining the performance of high-throughput assays relative to low-throughput analogs.

Predictive Model Building

When the concept of AOPs was initially proposed by Ankley et al. (2010), the authors noted the potential application of AOP frameworks for integrating mechanistic data with conventional animal-based studies to build predictive models. To be considered credible for use in regulatory decision making, predictive models must be built on a sound mechanistic foundation of the toxicological process (Wittwehr et al. 2017). This requirement is consistent with the underpinnings of the AOP conceptual framework and the reliance on defined relationships between an MIE and downstream key events, relationships that have been well established for the estrogen, androgen, and thyroid pathways. The interest in applying AOPs to regulatory scenarios and the coincident availability of HTS tools can promote a shift in chemical safety assessments from direct observations of chemical effects in animals to predictive models based on an understanding derived from mechanistic data from thousands of chemicals.

Alternative approaches/validation. Tens of thousands of registered chemicals must be screened for effects on human health and wildlife, and thousands of new chemicals are being developed every year. For all intents and purposes, chemical safety screening and the associated regulatory decision making continue to be dominated by mammalian-based *in vivo* testing, typically in a rodent model. The expense and time of continued reliance on animal testing is not a practical approach for effective safety screening, and new tactics are needed to close the gap between the number of chemicals in use and the number of chemicals assessed to date. Using computational and high-throughput screening alternatives to traditional toxicological methods

requires that new methods are appropriately interrogated to establish the soundness of the data produced (i.e., validation). One objective of traditional validation studies is to demonstrate method transferability, ensuring that any naïve lab can conduct the test and achieve satisfactory results (OECD 2005). Even for relatively simple *in vitro* test methods, interlaboratory "ring trials" may take years to complete and rely on relatively few chemicals. This aspect of validation is often not appropriate for HTS methods because assays are developed and conducted in one of only a few suitably equipped laboratories; thus, method transfer may not be a consideration.

High-throughput methods are amenable to a performance-based approach to validation, which determines the performance of proposed methods against a set of reference chemicals that are active (or inactive) over a range of potencies. For each molecular target, candidate reference chemicals can be identified and ideally are independent of the specific assay method used to identify the chemical activity. For example, active/inactive reference estrogen agonists may be identified from ER binding, ER transactivation, cell proliferation, or ER cofactor recruitment assays. Chemicals that are active in more than one type of assay can more reliably be considered "reference chemicals" for a biological effect and will reduce the possibility of including erroneous "reference" chemicals that interfere with the specific assay technology (e.g., chemophores, cytotoxic chemicals, assay interferences) rather than interact directly with the molecular target. High-throughput methods are uniquely suited for performance-based validation because the applicability domain and performance of the assay can be defined for a relatively large set of structurally diverse reference chemicals that span a wide range of potencies. For models with demonstrated performance against a large and robust set of reference chemicals, poor performance may also be informative. As relationships between endocrine mechanisms and downstream key events become better understood, poor prediction of a sequence of biological events may indicate chemical classes that are acting through other modes of action or are not adequately addressed by existing assays.

Simple predictive model: ER agonism. The ToxCast™ and Tox21 programs include high-throughput analogs of Tier 1 *in vitro* ER assays (e.g., ER-binding and ER transactivation assays) in addition to other HTS assays that measure ER signaling using a variety of different assay technologies and cell types. Concentration–response data from 18 ER HTS assays were integrated into an ER model, the output of which provides a score of potential ER agonist and antagonist activity, chemical potency, and a measure of assay-specific false positive activity of each chemical run in ToxCast™ (Judson et al. 2015). The variety of assay technologies and the redundancy of the 18 ER assays represent substantial benefits compared with the two Tier 1 EDSP *in vitro* ER assays.

To examine the performance of the ER model relative to the existing Tier 1 ER assays, the model scores were determined for *in vitro* ER reference chemicals identified by the Interagency Coordinating Committee on the Validation of Alternative Test Methods (ICCVAM; http://ntp.niehs.nih.gov/pubhealth/evalatm/iccvam/test-method-evaluations/endocrine-disruptors/in-vitro-assay-review/brd/index.html) and OECD (2012) for the express purpose of validating novel *in vitro* assays. Forty ER agonist reference chemicals with reproducible *in vitro* assay results included 28 agonists of differing potencies indicated by a range in half-maximal activity concentration (AC_{50} values and 12 inactive chemicals (Judson et al. 2015). The consensus list of reference chemicals was independent of assay type and for this reason, more likely to be "true," biologically relevant reference chemicals. The ER model predicted the activity of *in vitro* reference

chemicals with an overall accuracy of 93% and a false-negative rate of 7% (Browne et al. 2015).

The ability of the ER model to predict chemicals that were active/inactive in the *in vivo* rodent uterotrophic assay was also evaluated. Chemicals that act as *in vivo* estrogen agonists were identified from a systematic review of uterotrophic studies published in scientific journals, and those that were methodologically consistent with the EDSP Tier 1 guideline were regarded as "guideline-like" (Kleinstreuer et al. 2016a). Guideline-like uterotrophic studies were identified for 103 chemicals, and experimental details including chemical, dose, and uterine weight were extracted into a database (http://ntp.niehs.nih.gov/pubhealth/evalatm/tox21-support/endocrine-disruptors/edhts.html). Of the 103 chemicals with guideline-like studies, 43 chemicals had consistent ER agonist activities indicated by increased uterine weight (or a lack thereof) in two or more independent studies and were considered *in vivo* reference chemicals (Browne et al. 2015). The *in vivo* reference chemicals were then used to evaluate the ER model predictions of the *in vivo* response. Again, the ER model performance was excellent against *in vivo* reference chemicals, with an accuracy of 86% and a false-negative rate of 3% (Browne et al. 2015).

Based on the performance of the ER model against the 40 *in vitro* reference chemicals and 43 of the *in vivo* ER agonist reference chemicals (65 unique chemicals), the U.S. EPA published a *Federal Register* Notice stating the intention of the agency to accept computational tools and predictive models as alternative data for the current EDSP Tier 1 ER binding, ERTA, and rodent uterotrophic screening assays (U.S. EPA 2015b). The performance-based validation approach used to evaluate the ER model predictions against both *in vitro* and *in vivo* assays relies on presumptive relationships between the MIE (i.e., ER binding) and changes at the cellular (i.e., ERTA) and organ (i.e., change in uterine weight; Figure 5) levels consistent with the organization and interpretation of the EDSP Tier 1 screening battery data.

In addition to 18 ER assays, ToxCast™ and Tox21 include high-throughput alternatives to all EDSP Tier 1 *in vitro* assays. A similar model for androgen receptor (AR) interactions was developed based on 11 *in vitro* HTS assays, and AR model accuracy and precision were >95% against agonist and antagonist reference chemicals (Kleinstreuer et al. 2016b). The Tier 1 *in vitro* assays (ER, AR, aromatase inhibition, and steroidogenesis) can each be possible MIEs for endocrine-active chemicals or early key events altered in toxicity pathways (Figure 1). Efforts are presently underway to identify reference chemicals for each remaining MIE and to evaluate the performance of the HTS assays and compare results for chemicals run in high-throughput assays with corresponding low-throughput EDSP Tier 1 *in vitro* assays. The expectation is that HTS data will provide a suitable alternative for existing Tier 1 *in vitro* assays and, as in the uterotrophic assay example, may predict downstream *in vivo* key events.

Complex Predictive Models. High-throughput nonanimal alternative methods can be used to rapidly screen thousands of chemicals and are poised to be critical tools in modernizing chemical safety evaluation. However, most available high-throughput methods measure a specific mechanism or key event and are therefore incapable of recapitulating all possible physiological responses in an animal (Coady et al. 2017). Rather than predicting the full spectrum of possible *in vivo* responses, a more attainable near-term goal may be to predict the specific *in vivo* end points that inform chemical regulatory decisions.

AOP frameworks can help to reduce the overwhelming complexity of animal physiology to isolate end points evaluated in regulatory decision making, can determine how well alternative methods predict *in vivo* apical responses, and can be used to develop integrated testing strategies (OECD 2017). Integrated testing strategies can be used for interim determinations of testing needs, can help to determine the specific data required to arrive at regulatory conclusions, and may include rule-based decisions directing progressively higher tiered testing (Vinken 2013; OECD 2017). Ideally, testing strategies are developed around an understanding of mechanisms of toxicity and the sequelae of downstream responses (Tollefsen et al. 2014) and can integrate results derived from a combination of methods (e.g., *in silico*, *in vitro*, *in vivo* approaches; OECD 2017). As discussed previously, to be useful for regulatory decision making, these AOPs do not have to exhaustively cover all key events from MIE to adverse outcome, nor are the end points considered in EDSP Tier 1 screening assays or in other standardized methods used to inform regulatory decisions.

The EDSP approach for screening and testing chemicals for potential disruption is a testing strategy built around pathway frameworks. For example, WoE evaluations of List 1 Tier 1 screening were used to direct Tier 2 testing requirements (https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-tier-1-assessments). Data from the ToxCast™ ER model are now used to direct additional screening and testing requirements for the estrogen agonist pathway, and as additional
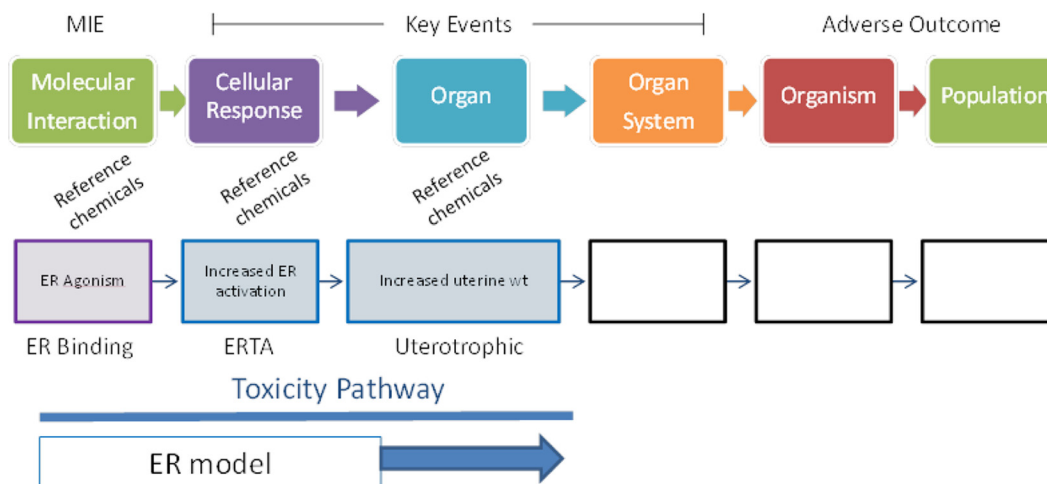

Figure 5. Validation of an estrogen receptor (ER) agonism toxicity pathway model based on ToxCast™/Tox21 high-throughput screening (HTS) data requires a robust set of reference chemicals for the relevant molecular initiating event (MIE) and key events (Browne et al. 2015).

Table 1. U.S. EPA Endocrine Disruptor Screening Program (EDSP) tier 1 screening battery assays and tier 2 testing assays, high-throughput screening (HTS) assays, and predictive model alternatives (U.S. EPA, 2015a; Kleinstreuer et al. 2016b).

Level	EDSP assays	HTS assays and predictive model alternatives
Tier 1 screening battery	Estrogen receptor (ER) binding	ER Model
Tier 1 screening battery	Estrogen receptor transactivation (ERTA)	ER Model
Tier 1 screening battery	Uterotrophic	ER Model
Tier 1 screening battery	Androgen receptor (AR) binding	AR Model
Tier 1 screening battery	Hershberger	AR Model
Tier 1 screening battery	Aromatase inhibition	STR
Tier 1 screening battery	Steroidogenesis (STR)	STR
Tier 1 screening battery	Female rat pubertal	ER, STR, THY
Tier 1 screening battery	Male rat pubertal	AR, STR, THY
Tier 1 screening battery	Fish short term reproduction	ER, AR, STR
Tier 1 screening battery	Amphibian metamorphosis	THY
Tier 2 definitive test	Rat 2-gen/EOGRT	ER, AR, STR, THY
Tier 2 definitive test	MEOGRT	ER, AR, STR
Tier 2 definitive test	LAGDA	THY
Tier 2 definitive test	Japanese Quail 2-gen	ER, AR, STR, THY

Note: AR, androgen receptor; EOGRT, extended one-generation reproductive toxicity; ER, estrogen receptor; LGDA, larval amphibian growth and development assay; MEOGRT, Medaka extended one-generation reproductive test; STR, steroidogenesis; THY, thyroid. Bold type indicates totals. For whole-animal *in vivo* assays (e.g., female rat pubertal assay), several predictive models may be needed in an integrated approach to testing and assessment (IATA).
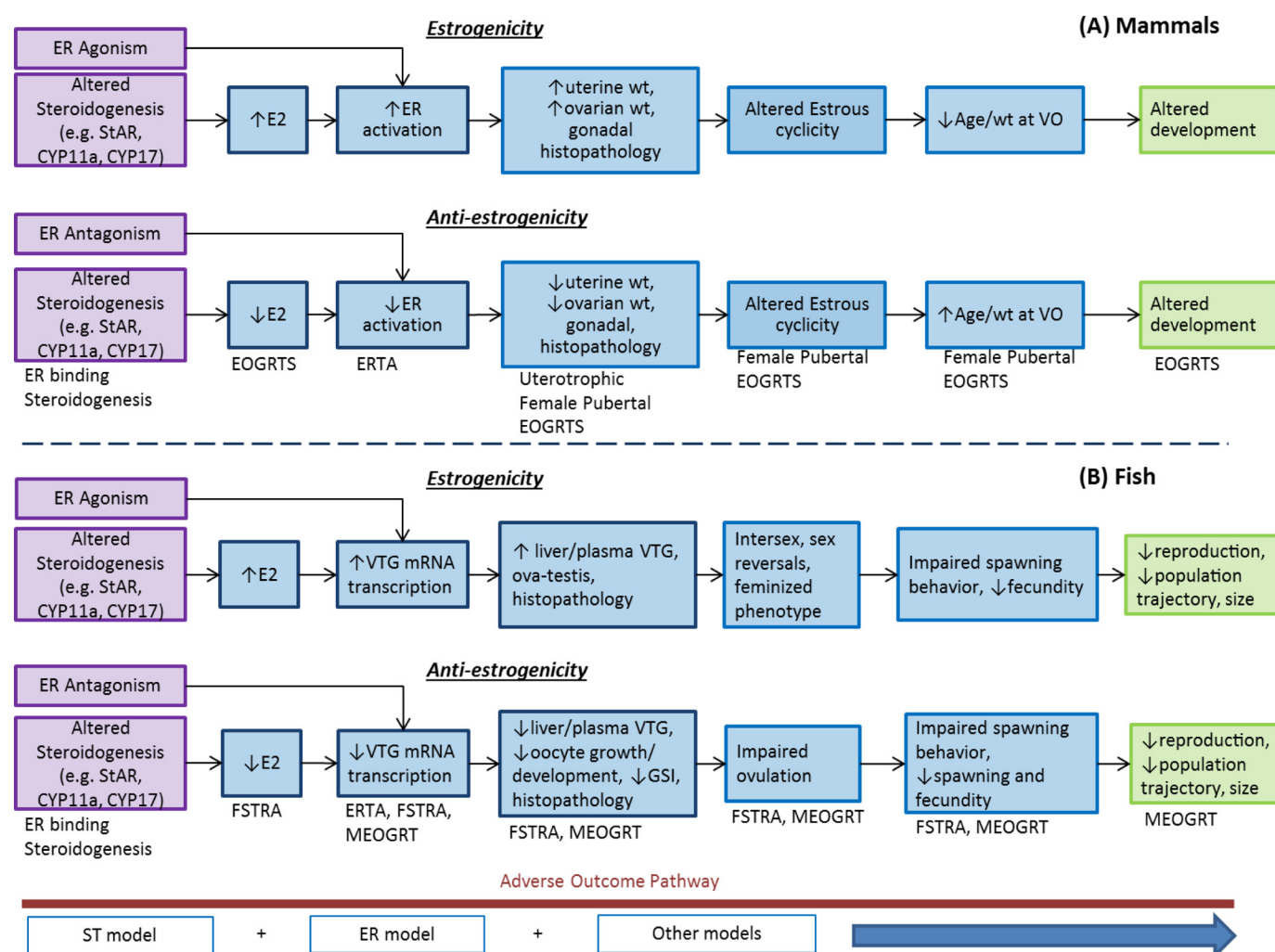
Figure 6. Adverse outcome pathway (AOP) networks for estrogenicity and antiestrogenicity in (*A*) mammals and (*B*) fish based on integrated estrogen receptor (ER) agonist and steroidogenesis modeling outputs along with Endocrine Disruptor Screening Program (EDSP) Tier 1 and Tier 2 assays. Additional modeling may be integrated into the AOP network to further refine the model and toxicity pathway/AOP linkages. CYP11a, cytochrome P450 11a (cholesterol side-change cleavage protein); CYP17, cytochrome P450 17 (e.g., 17α-hydroxylase); E2, 17β-estradiol; EOGRTS, extended one-generation reproductive toxicity study; FSTRA, fish short-term reproductive assay; MEOGRT, Medaka extended one-generation reproductive test; StAR, steroidogenic acute regulatory protein; VO, vaginal opening; VTG, vitellogenin.

alternative methods are integrated in screening other endocrine pathways, the results from these approaches are expected to influence additional screening and testing requirements.

Developing alternative methods for predicting apical *in vivo* responses to environmental chemicals will likely evolve from multiple predictive models that can be integrated in increasingly complex approaches to model organismal responses. For example, using alternative methods to predict endocrine effects of chemicals that alter reproductive development in the rat female pubertal assay will likely necessitate the development of models for estrogen, steroidogenesis, and thyroid pathway effects (Table 1; Figure 6). These more simple predictive models can be validated separately against an appropriate set of reference chemicals for *in vitro* and *in vivo* responses, and then models can be considered together in an integrated testing strategy to predict key event end points (e.g., OECD 2016) and complex biological responses (Joworska and Hoffmann 2010). AOPs and AOP networks provide a systematic approach for interpreting the biological relevance of alternative methods, evaluating the utility of these methods for making regulatory decisions, identifying additional data needs, and determining under what circumstances increasingly resource-intensive assays might be needed to reduce uncertainty in chemical safety assessments (Allen et al. 2014; Tollefsen et al. 2014; Burden et al. 2015; Patlewicz et al. 2015; OECD 2017).

As more data become available for retrospective analyses, AOPs will continue to be used to build predictive models and can include quantitative AOPs to predict the magnitude of downstream effects and provide dose–response relationships (Villeneuve et al. 2014a). Quantitative AOPs require a detailed understanding of not only the key events in the AOP but also the temporality and magnitude of the key event relationships, and for this reason, they may take years to develop and validate. Once developed, quantitative AOPs can provide alternatives to the model organism or population responses currently needed for risk assessment (Groh et al. 2015), and quantitative predictions determined for a specific chemical can be extrapolated to chemicals with shared modes of action (e.g., Conolly et al. 2017), which may substantiate the resource investment for critical regulatory end points or susceptible wildlife populations.

Moreover, AOP networks can be used to define chemical categories and to indicate points of convergence in biological targets in multiple pathways for which assays could be designed. Points of convergence or common biological targets in AOP networks may be considered "tipping points" or candidate biomarkers. Broadly defined, biomarkers are measurable biological responses (cellular, biochemical, physiological) of a cell or organism that can be used to monitor exposure to and effects of a chemical (Biomarkers Definitions Working Group 2001; Robb et al. 2016). In the EDSP, for example, changes in vitellogenin (VTG) mRNA and protein levels that lead to potential declines in reproductive success and population trajectories have been used as biomarkers of chemical-related estrogenic activity in oviparous vertebrates. Similarly, chemically induced changes in levels of circulating hormone and histopathological measurements in rodents and other taxa may serve as biomarkers indicative of adverse outcomes characterized by altered development and reproductive function (e.g., delayed/accelerated puberty). Identifying tipping points in endocrine AOPs can help distinguish between biological responses that are adaptive (expected to be early events in the AOP) and responses predictive of adversity or toxicity (expected to be downstream in the AOP). With AOP development, validation, and application, new *in silico* and *in vitro* methods targeting key events could provide sufficient information for hazard and risk assessments with little to no *in vivo* testing (e.g., MacKay et al. 2013).

Although there are many advantages to using AOPs in the EDSP and other regulatory contexts, there are recognized limitations of the AOP framework approach. First, AOPs are chemical-agnostic (Villeneuve et al. 2014a) and therefore are not directly relevant to the hazard identification of a specific chemical. Nonetheless, one would expect chemicals with the same mode of action to have similar patterns of biological responses across assays; therefore, AOPs can be used in chemical categorization and read-across (OECD 2013, 2014). Further, adverse outcomes may be caused by unknown modes of action that are not targeted in the EDSP screening battery or HTS methods and so may not be captured in a particular data set and AOP. AOPs are highly reductive, and key event relationships may be correlative rather than causal. As a result, there is a possibility that an observed response may be attributed to an endocrine mechanism, when in fact it is due to a different mode of action; this may carry substantial regulatory implications. This possibility is a particular concern for chemical exposures that result in systemic toxicity, which may confound endocrine effects. Efforts have been made to distinguish between nonspecific *in vitro* activity (Judson et al. 2015) and the effects of systemic toxicity in EDSP Tier 1 assays (https://www.epa.gov/sites/production/files/2015-06/documents/052113minutes.pdf), but care must be taken to consider these effects in endocrine AOP development. An increasingly diverse set of HTS assays provide coverage of nonendocrine pathways. Other *in vitro* pathway activities, respective potencies, and *in vivo* data may help to discriminate between mechanisms of toxicity. Additionally, AOP constructs do not always account for compensatory mechanisms or modulating factors that may influence dose–response and apical outcomes. An example of an AOP including the inhibitory feedback of progesterone on estrogenic responses in the uterotrophic assay (Simon et al. 2014) has been published, but this is more the exception than the rule for many current AOPs. Thus far, applications of AOP concepts in endocrine screening have mostly focused on using AOPs as a systematic organizing construct and as a way to integrate new toxicological methods in a testing strategy developed 20 y ago. The U.S. EPA has modified the requirements for EDSP Tier 1 battery screening based on the availability of ToxCast™ ER model data, and with growing understanding of underlying biology, the availability of new assay technologies, and appropriate demonstration of the performance of other predictive models, the EDSP Tier 1 screening battery is likely to change to include the evolving science.

Conclusions

The use of toxicity pathways and AOP frameworks offers the potential to improve the understanding and prediction of endocrine disruption. AOPs are an organizing framework for multiple types of disparate data measured at different levels of biological organization and can be used to better evaluate all available data in a WoE evaluation of a chemical's potential endocrine activity or other biological activities relevant to adverse outcomes. In this paper, we have described the use of AOPs built around existing regulatory end points that do not include all key events and thus have gaps in the key events and key event relationships. Nonetheless, pathway frameworks are very useful tools for attributing a general response (e.g., decreased fertility) to a specific endocrine mode of action (e.g., estrogenic signaling in males) or for identifying critical knowledge gaps that prevent determination of a mode of action to a chemical. In these examples, the use of AOPs helps to *a*) include other data (than end points measured in guideline studies) that may support or refute the putative relationship between the existing end points, *b*) better evaluate the overall consistency of the responses and the possible alternative modes of action, and *c*) identify key data gaps that are needed to build plausible

relationships between mechanisms and apical responses used for risk assessment.

Pathway frameworks also facilitate the evaluation of alternatives to methods currently in practice and as possible replacements for *in vivo* animal testing. A pathway-based approach helps establish confidence in the *in vitro* prediction of *in vivo* results, identifies data gaps, and guides further research. Improved mechanistic understanding facilitates development of alternative tests, aids extrapolation across species by facilitating comparative analysis of toxicity information across species, and focuses testing on key targets associated with AOPs or AOP networks. In the EDSP context, AOP networks model the interactions between multiple endocrine pathways, identify possible points of convergence, and may identify potential biomarkers around which assays and testing strategies may be developed. Strengthening the linkage between activity and adverse effects in individuals or populations would provide the basis for more meaningful inclusion of endocrine activity data into risk assessments. The U.S. EPA's EDSP is a unique regulatory program that was developed around a mode of action framework and thus is a logical demonstration of how AOP concepts can be used in regulatory science. The EDSP provides examples of how AOP concepts can be used to interpret specific end points from multiple independent guideline studies that may be the only data available for regulatory decision making. In addition, AOPs support the development of alternative and strategic testing methods. In the future, AOPs will likely be used to support increasingly sophisticated models to predict complex *in vivo* end points that can continue to reduce animal use and can increase the rate and efficacy of chemical screening.

Acknowledgments

References

Allen TEH, Goodman JM, Gutsell S, Russell PJ. 2014. Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. Chem Res Toxicol 27(12):2100–2112, PMID: 25354311, https://doi.org/10.1021/tx500345j.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, et al. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730–741, PMID: 20821501, https://doi.org/10.1002/etc.34.

Ankley GT, LaLone CA, Gray LE, Villeneuve DL, Hornung MV. 2016. Evaluation of the scientific underpinnings for identifying estrogenic chemicals in nonmammalian taxa using mammalian test systems. Environ Toxicol Chem 35(11):2806–2816, PMID: 27074246, https://doi.org/10.1002/etc.3456.

Ankley GT, Villeneuve DL. 2015. Temporal changes in biological responses and uncertainty in assessing risk of endocrine-disrupting chemicals: insights from intensive time-course studies with fish. Toxicol Sci 144(2):259–275, PMID: 25564424, https://doi.org/10.1093/toxsci/kfu320.

Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin Pharmacol Ther 69(3):89–95, PMID: 11240971, https://doi.org/10.1067/mcp.2001.113989.

Becker RA, Ankley GT, Edwards SW, Kennedy SW, Linkov I, Meek B, et al. 2015. Increasing scientific confidence in adverse outcome pathways: Application of tailored Bradford-Hill considerations for evaluating weight of evidence. Reg Toxicol Pharmacol 72(3):514–537, PMID: 25863193, https://doi.org/10.1016/j.yrtph.2015.04.004.

Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS. 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. Environ Sci Technol 49(14):8804–8814, PMID: 26066997, https://doi.org/10.1021/acs.est.5b02641.

Burden N, Sewell F, Andersen ME, Boobis A, Chipman JK, Cronin MT, et al. 2015. Adverse outcome pathways can drive non-animal approaches for safety assessment. J Appl Toxicol 35(9):971–975, PMID: 25943792, https://doi.org/10.1002/jat.3165.

Coady KK, Biever RC, Denslow ND, Gross M, Guiney PD, Holbech H, et al. 2017. Current limitations and recommendations to improve testing for the environmental assessment of endocrine active substances. Integr Environ Assess Manag. 13(2):302–316, PMID: 27791330, https://doi.org/10.1002/ieam.1862.

Colborn T, Clement C. 1992. Chemically induced alterations in sexual and functional development: The wildlife/human connection. In: *Advances in Modern Environmental Toxicology*, Vol. 21. Mehlman MA, ed. Princeton, NJ:Princeton Scientific Publishing Co., Inc.

Conolly RB, Ankley GT, Cheng WY, Mayo ML, Miller DH, Perkins EJ, Villeneuve DL, Watanabe KH. 2017. Quantitative adverse outcome pathways and their applciation to predictive toxicology. Environ Sci Technol 51(8):4661–4672, PMID: 28355063, https://doi.org/10.1021/acs.est.6b06230.

EDSTAC (Endocrine Disruptor Screening and Testing Advisory Committee). 1998. "Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) Final Report." Washington, DC: U.S. Environmental Protection Agency. http://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-and-testing-advisory-committee-edstac-final [accessed 21 June 2017].

Filer D, Patisaul HB, Schug T, Reif D, Thayer K. 2014. Test driving ToxCast: Endocrine profiling for 1858 chemicals included in phase II. Curr Opin Pharmacol 19:145–152, PMID: 25460227, https://doi.org/10.1016/j.coph.2014.09.021.

Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, et al. 2015. EDC-2: The Endocrine Society's second scientific statement on endocrine-disrupting chemicals. Endocr Rev 36(6):E1–E150, PMID: 26544531, https://doi.org/10.1210/er.2015-1010.

Groh KJ, Carvalho RN, Chipman JK, Denslow ND, Halder M, Murphy CA, et al. 2015. Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology. Chemosphere 120:764–777, PMID: 25439131, https://doi.org/10.1016/j.chemosphere.2014.09.068.

Hallinger DR, Murr AS, Buckalew AR, Simmons SO, Stoker TE, Laws SC. 2017. Development of a screening approach to detect thyroid disrupting chemicals that inhibit the human sodium iodide symporter (NIS). Toxicol In Vitro 40:66–78, PMID: 27979590, https://doi.org/10.1016/j.tiv.2016.12.006.

Hotchkiss AK, Rider CV, Blystone CR, Wilson VS, Hartig PC, Ankley GT, et al. 2008. Fifteen years after "Wingspread" – Environmental endocrine disrupters and human and wildlife health: Where we are today and where we need to go. Toxicol Sci 105(2):235–259, PMID: 18281716, https://doi.org/10.1093/toxsci/kfn030.

Joworska J, Hoffman S. 2010. Integrated Testing Strategy (ITS) - Opportunities to better use existing data and guide future testing in toxicology. ALTEX 27(4):231–242, PMID: 21240466.

Judson RS, Magpantay FM, Chickarmane V, Haskell C, Tania N, Taylor J, et al. 2015. Integrated model of chemical perturbations of a biological pathway using 18 *in vitro* high-throughput screening assays for the estrogen receptor. Toxicol Sci 148(1):137–154, PMID: 26272952, https://doi.org/10.1093/toxsci/kfv168.

Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, et al. 1996. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: A report of the U.S. EPA-sponsored workshop. Environ Health Perspect 104:715–740, PMID: 8880000, https://doi.org/10.2307/3432708.

Kleinstreuer NC, Ceger PC, Allen DG, Strickland J, Chang X, Hamm JT, et al. 2016a. A curated database of rodent uterotrophic bioactivity. Environ Health Perspect 124(5):556–562, PMID: 26431337, https://doi.org/10.1289/ehp.1510183.

Kleinstreuer NC, Ceger P, Watt ED, Martin M, Houck K, Browne P, et al. 2016b. Development and validation of a computational model for androgen receptor activity. Chem Res Toxicol 30(4):946–964, PMID: 27933809, https://doi.org/10.1021/acs.chemrestox.6b00347.

Knapen D, Vergauwen L, Villeneuve DL, Ankley GT. 2015. The potential of AOP networks for reproductive and developmental toxicity assay development. Reprod Toxicol 56:52–55, PMID: 25889759, https://doi.org/10.1016/j.reprotox.2015.04.003.

Kramer VJ, Etterson MA, Hecker M, Murphy CA, Roesijadi G, Spade DJ, et al. 2011. Adverse outcome pathways and ecological risk assessment bridging to population-level effects. Environ Toxicol Chem 30(1):64–76, PMID: 20963853, https://doi.org/10.1002/etc.375.

LaLone CA, Villeneuve DL, Burgoon LD, Russom CL, Helgen HW, Berninger JP, et al. 2013. Molecular target sequence similarity as a basis for species extrapolation to assess the ecological risk fo chemicals with known modes of action. Aquat Toxicol 144-145:141–154, PMID: 24177217, https://doi.org/10.1016/j.aquatox.2013.09.004.

LaLone CA, Villeneuve DL, Lyons D, Helgen HW, Robinson SL, Swintek JA, et al. 2016. Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A web-based tool for addressing the challenges of cross-species extrapolation of chemical toxicity. Toxicol Sci 153:228–245, PMID: 27370413, https://doi.org/10.1093/toxsci/kfw119.

LaLone CA, Villeneuve DL, Wu-Smart J, Milsk RY, Sappington K, Gerver KV, et al. 2017. Weight of evidence evaluation of a network of adverse outcome pathways linking activation of the nicotinic acetylchloine receptor in honey bees to colony death. Sci Total Environ 584–585:751–775, PMID: 28126277, https://doi.org/10.1016/j.scitotenv.2017.01.113.

MacKay C, Davies M, Summerfield V, Maxwell G. 2013. From pathways to people: Applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. ALTEX 30(4):473–486, PMID: 24173169, https://doi.org/10.14573/altex.2013.4.473.

Makris SL, Scott CS, Fox J, Knudsen TB, Hotchkiss AK, Arzuaga X, et al. 2016. A systematic evaluation of the potential effects of trichloroethylene exposure on cardiac development. Reprod Toxicol 65:321–358, PMID: 27575429, https://doi.org/10.1016/j.reprotox.2016.08.014.

Nohynek GJ, Borgert CJ, Dietrich D, Rozman KK. 2013. Endocrine disruption: Fact or urban legend?. Toxicol Lett 223(3):295–305, PMID: 24177261, https://doi.org/10.1016/j.toxlet.2013.10.022.

NAS (National Academies of Sciences, Engineering, and Medicine). 2017. *Using 21st Century Science to Improve Risk-Related Evaluations*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24635.

NRC (National Research Council). 2007. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. http://www.nap.edu/openbook.php?record_id=11970. Washington, DC:National Research Council of the National Academies, National Academies Press.

OECD (Organisation for Economic Co-operation and Development). 2005. OECD series on testing and assessment number 34. Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2005)14 [accessed 21 March 2017].

OECD. 2012. OECD guideline for the testing of chemicals no. 457: BG1Luc estrogen receptor transactivation test method for identifying estrogen receptor agonists and antagonists. http://www.oecd-ilibrary.org/environment/test-no-457-bg1luc-estrogen-receptor-transactivation-test-method-for-identifying-estrogen-receptor-agonists-and-antagonists_9789264185395-en [accessed 3 May 2016].

OECD. 2013. OECD series on testing and assessment no. 184. Guidance document on developing and assessing adverse outcome pathways. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&docLanguage=en [accessed 6 June 2017].

OECD. 2014. OECD series on testing and assessment no. 194. Guidance on grouping chemicals. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono%282014%294&doclanguage=en [accessed 7 June 2017].

OECD. 2016. OECD series on testing and assessment no. 256. Annex 1: Case studies to the guidance document on the reporting of defined approaches and individual information sources to be used within integrated approaches to testing and assessment (IATA) for Skin Sensitisation. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono%282016%2929/ann1&doclanguage=en [accessed 7 June 2017].

OECD. 2017. OECD series on testing and assessment no. 260. Guidance document for the use of adverse outcome pathways in developing integrated approaches to testing and assessment (IATA). http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)67&doclanguage=en [accessed 21 June 2017].

Patlewicz G, Simon TW, Rowlands JC, Budinsky RA, Becker RA. 2015. Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. Reg Toxicol Pharmacol 71(3):463–477, PMID: 25707856, https://doi.org/10.1016/j.yrtph.2015.02.011.

Paul Friedman K, Watt ED, Hornung MW, Hedge JM, Judson RS, Crofton KM, et al. 2016. Tiered high-throughput screening approach to identify thyroperoxidase inhibitors within the ToxCast phase I and II chemical libraries. Toxicol Sci 151(1):160–180, PMID: 26884060, https://doi.org/10.1093/toxsci/kfw034.

Reif DM, Martin MT, Tan SW, Houck KA, Judson RS, Richard AM, et al. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ Health Perspect 118(12):1714–1720, PMID: 20826373, https://doi.org/10.1289/ehp.1002180.

Robb MA, McInnes PM, Califf RM. 2016. Biomarkers and surrogate endpoints: Developing common terminology and definitions. JAMA 315(11):1107–1108, PMID: 26978201, https://doi.org/10.1001/jama.2016.2240.

Rotroff DM, Dix DJ, Houck KA, Knudsen TB, Martin MT, McLaurin KW, et al. 2013. Using *in vitro* high throughput screening assays to identify potential endocrine-disrupting chemicals. Environ Health Perspect 121(1):7–14, PMID: 23052129, https://doi.org/10.1289/ehp.1205065.

Simon TW, Simons SS Jr, Preston RJ, Boobis AR, Cohen SM, Doerrer NG, et al. 2014. The use of mode of action information in risk assessment: Quantitative key events/dose-response framework for modeling the dose-response for key events. Crit Rev Toxicol 44(suppl 3):17–43, PMID: 25070415, https://doi.org/10.3109/10408444.2014.931925.

Soto AM, Sonnenschein C. 2010. Environmental causes of cancer: Endocrine disruptors as carcinogens. Nat Rev Endocrinol 6(7):364–371, PMID: 20498677, https://doi.org/10.1038/nrendo.2010.87.

Teeguarden JG, Tan YM, Edwards SW, Leonard JA, Anderson KA, Corley RA, et al. 2016. Completing the link between exposure science and toxicology for improved environmental health decision making: The aggregate exposure pathway framework. Environ Sci Technol 50(9):4579–4586, PMID: 26759916, https://doi.org/10.1021/acs.est.5b05311.

Tollefsen KE, Scholz S, Cronin MT, Edwards SW, de Knecht J, Crofton K, et al. 2014. Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). Reg Toxicol Pharmacol 70(3):629–640, PMID: 25261300, https://doi.org/10.1016/j.yrtph.2014.09.009.

U.S. Congress. 1996a. Food Quality Protection Act of 1996. Pub L No. 104–170 (104th Congress, 3 August 1996). https://www.gpo.gov/fdsys/pkg/PLAW-104publ170/pdf/PLAW-104publ170.pdf [accessed 2 May 2016].

U.S. Congress. 1996b. The Safe Drinking Water Act Amendments of 1996. Pub L No. 104–182. (104th Congress, 6 August 1996). http://www.gpo.gov/fdsys/pkg/PLAW-104publ182/pdf/PLAW-104publ182.pdf [accessed 2 May 2016].

U.S. EPA (U.S. Environmental Protection Agency). 2009. Overview of the first list of chemicals for Tier 1 screening under the Endocrine Disruptor Screening Program. https://www.epa.gov/endocrine-disruption/overview-first-list-chemicals-tier-1-screening-under-endocrine-disruptor [accessed 2 May 2016].

U.S. EPA. 2011. Weight-of-evidence: Evaluating results of EDSP Tier 1 screening to identify the need for Tier 2 testing. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2010-0877-0021 [accessed 8 March 2016].

U.S. EPA. 2012. U.S. Environmental Protection Agency Endocrine Disruptor Screening Program: Universe of chemicals and general validation principles. http://www.epa.gov/sites/production/files/2015-07/documents/edsp_chemical_universe_and_general_validations_white_paper_11_12.pdf [accessed 8 March 2016].

U.S. EPA. 2015a. Endocrine Disruptor Screening Program Tier 1 assessments. https://www.epa.gov/ingredients-used-pesticide-products/endocrine-disruptor-screening-program-tier-1-assessments [accessed 2 May 2016].

U.S. EPA. 2015b. Use of high throughput assays and computational tools; Endocrine Disruptor Screening Program; Notice of availability and opportunity for comment. Fed Reg 80(118):35350–35355. https://www.federalregister.gov/articles/2015/06/19/2015-15182/use-of-high-throughput-assays-and-computational-tools-endocrine-disruptor-screening-program-notice [accessed 7 March 2016].

U.S. EPA. 2015c. Human health risk assessment: Strategic research action plan 2016-2019. https://www.epa.gov/sites/production/files/2015-10/documents/strap_2016_hhra_508.pdf [accessed 14 March 2017].

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, et al. 2014a. Adverse outcome pathway (AOP) development I: Strategies and principles. Toxicol Sci 142(2):312–320, PMID: 25466378, https://doi.org/10.1093/toxsci/kfu199.

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, et al. 2014b. Adverse outcome pathway development II: Best practices. Toxicol Sci 142(2):321–330, PMID: 25466379, https://doi.org/10.1093/toxsci/kfu200.

Vinken M. 2013. The adverse outcome pathway concept: A pragmatic tool in toxicology. Toxicology 312:158–165, PMID: 23978457, https://doi.org/10.1016/j.tox.2013.08.011.

Whelan M, Andersen M. 2013. "Toxicity Pathways - from concepts to application in chemical safety assessment." Luxembourg:Publications Office of the European Union. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC86467/lb-na-26389-en-n%20.pdf [accessed 2 May 2016].

WHO (World Health Organization). 2002. *Global assessment of the state-of-the-science of endocrine disruptors*. http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/ [accessed 2 September 2016].

WHO/UNEP (United Nations Environmental Programme). 2012. *State of the science of endocrine disrupting chemicals - 2012*. http://www.who.int/ceh/publications/endocrine/en/ [accessed 2 September 2016].

Wittwehr C, Aladjov H, Ankley G, Byrne HJ, de Knecht J, Heinzle E, et al. 2017. How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology. Toxicol Sci 155(2):326–336, PMID: 27994170, https://doi.org/10.1093/toxsci/kfw207.